Pattern Recognition Letters 000 (2015) 1–8

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface ☆

Hubert Cecotti*

*School of Computing and Intelligent Systems, University of Ulster, Londonderry, Northern Ireland, UK*

ARTICLE INFO

*Article history:*
Available online xxxx

*Keywords:*
Event-related potentials (ERP)
Brain-computer interface
Shift invariant detection
Feature extraction

ABSTRACT

Non-invasive brain-computer interface (BCI) is a relatively new type of human-computer interaction. BCIs that are based on the detection of event-related potentials (ERPs) are usually synchronous. They require the knowledge of the stimulus onsets that evoke ERPs, which is time locked to the presence of a potentially relevant stimulus. The detection of ERPs like the P300 has been successfully used in BCI thanks to the oddball paradigm. The time locked detection is directly related to the synchronous aspect of a BCI. However, asynchronous detection is a critical issue in developing BCIs for real-life applications, where the machine should be able to detect the presence of an ERP independently from the knowledge of the stimulus onsets, or when wireless devices do not allow a precise knowledge of the stimulus onsets. Although the detection of single-trial ERP is already a challenge, when the stimulus onsets are well identified, we propose to investigate further the detection of single-trial ERP by considering different time locked stimuli. We propose and compare shift invariant ERP detection strategies on data from ten subjects obtained in a P300 speller experiment. With a shift invariant distance, we show that it is possible to obtain an AUC of 0.834 while allowing a jitter of ±40 ms. With inputs in the Fourier domain, the mean area under the ROC curves of 0.683 allowing a jitter of ±200 ms in the stimulus onsets. The results support the conclusion that ERP detection can be achieved without a precise knowledge of the stimulus onsets, and hence can be used with EEG amplifiers that do not allow a precise synchronization between the EEG signal and stimulus onsets.



## 1. Introduction

Brain-computer interface (BCI) is a type of human-computer interaction that is based on the detection of the user's brain responses. An event-related potential (ERP) is the measured brain response that is the direct result of an event such as the presentation of a visual stimulus [30], e.g., affective images [18,25]. In BCI, it typically represents the electrophysiological response (in microvolt) to a stimulus. The detection of event-related potentials in non-invasive brain-computer interfaces has been used for more than two decades [12]. The P300 speller has been one of the leading applications in which the detection of ERPs is achieved thanks to the oddball paradigm [31]. BCIs based on the oddball paradigm are typically synchronous as the ERP detection takes into account the stimulus onsets, i.e., each detection is time locked to the presentation of a stimulus. Due to the low signal-to-noise ratio of the EEG signal, the detection of ERPs requires efficient and robust pattern recognition techniques that can deal with the non-stationarity of the signal and the specificity of each subject.

While this type of BCI is popular, it remains a challenge to obtain a functioning synchronous BCI based on ERP detection outside of the laboratory. As an alternative, some early works have directly focused their attention on asynchronous BCI [33]. In an asynchronous protocol, the subject makes self-paced decisions, such that she/he decides voluntarily when to switch between mental tasks. An asynchronous system is usually decomposed into two parts. First, the system needs to detect if the subject is in an idle state or an action state. Second, the system needs to determine the type of action, i.e., the BCI command. In many asynchronous BCIs, the user produces a voluntary brain response, and the confidence value of the detection must reach a threshold in order to produce a command. If the confidence value does not reach the threshold, the user has to reproduce a particular and expected brain response until it is correctly detected. This step can be continuous like for the detection of steady states visual evoked potentials (SSVEP): the user keeps looking at a flickering visual stimulus until a response is produced [4], or for the detection of motor imagery [27]. In this case, the user can keep imagining a movement until a command is produced.

In BCI, a key asynchronous feature comes typically from temporal variability with which the user makes a decision. For instance, it is possible to have an asynchronous P300-BCI whereas the stream of visual stimuli stays synchronous. In such a case, one part of the

☆ This paper has been recommended for acceptance by G. Sanniti di Baja.

* Corresponding author. Tel.: +44 28 71675276.
*E-mail address:* hub20xx@hotmail.com, h.cecotti@ulster.ac.uk

http://dx.doi.org/10.1016/j.patrec.2015.01.015
0167-8655/

system detects when the user pays attention to the virtual keyboard, but the ERP detection stays synchronous. In [37], an asynchronous BCI that combines the P300 and SSVEP paradigms is proposed. The information transfer is accomplished using P300 ERP paradigm, and the control state detection is achieved using SSVEP, overlaid on the P300 based system. However, asynchronous features can also come from other aspects of the task. Contrary to the classical understanding of asynchronous BCIs, here we address the problem of asynchronicity, not when the subject takes self-paced decisions, but rather when the stream of visual stimuli is asynchronous. When the stream of visual stimuli is asynchronous, a visual stimulus can happen anytime. In this situation, the trigger corresponding to the event when a stimulus is presented to the user is unknown. This information is therefore missing for the classifier for performing single-trial detection. This problem is relevant, as it could have a critical impact for the user in some applications when an important stimulus is missed, and cannot be repeated. Besides, it would be possible to overcome the necessity of the stimulus onsets.

BCIs have been used for target detection by using rapid serial visual presentation (RSVP) tasks. An RSVP paradigm is a useful tool for researchers working on visual attention, allowing researchers to study the temporal characteristics of neural information processing [11]. In this paradigm, each image replaces the previous one at the same spatial location. Single-trial detection of ERPs has been addressed in several problems such as target detection by using RSVP tasks, e.g., the search of targets in satellite images [3,15,38], face recognition [46]. Target detection has also been the subject of a recent competition in an international machine learning workshop (MLSP) [17]. In these RSVP tasks, the knowledge of the stimulus onsets, i.e., when a stimulus is presented, is known both during training and the test phase. Thus, RSVP tasks are currently synchronous where a decision is given for the presentation of each visual stimulus. In this case, the ERP detection is time locked to the exact presentation of the different stimuli. Nonetheless, we have to know the stimulus onsets to create a ground truth for estimating the performance of a system that detects ERPs without a precise knowledge of the different visual stimuli presentation. Furthermore, ERP detection can be used for other scenarios of target detection, e.g., video, observation of the natural environment. Those scenarios prevent the use of current techniques due to the absence of precise information about the stimulus onsets (it is impossible to know when an event will happen).

During a typical RSVP task, the possible targets are presented to the user in a precise pace known by the detection procedure. In such a case, potential targets are already selected and presented to the user. When a subject has to detect a potential target in an environment, e.g., virtual reality, battlefield, it is not possible to present the target at a pre-defined moment: the presentation of the target can happen at any moment. For this reason, the possibility to detect a target at any time would benefit several applications, and extend the possible paradigms for ERP analysis. Moreover, it is the only practical case when single-trial detection can be truly justified. In current RSVP tasks such as large-scale image database triage [20,21], the images are already pre-processed before their presentation. Thus, nothing constraints the application to the presentation of previously seen images to the subject. If the images can be presented several times to the subject, then a detection strategy based on averaged trials could be applied. Although single-trial detection is the ultimate goal, its current need is mainly justified for speed reason in existing applications. Single-trial detection is definitively required when the stream of visual stimuli happens in real time, and each visual stimulus can only be processed one single time. When new stimuli are presented in real time, the repetition of a stimulus is not possible, as it would hide the presentation of stimulus at the current time.

The present work is motivated by applications that include real-time monitoring, and where events occur over time at unknown moments, and cannot be repeated. We propose the evaluation of the robustness of a BCI based on the detection of ERPs when the inputs are shifted in time. This evaluation will determine the type of jitter that can be conceded for an efficient detection. The analysis of performance for single-trial detection with shift invariant features in the Fourier domain is then evaluated. We show that such a strategy can benefit to single-trial detection when stimulus onsets are unknown or when an important jitter is present.

The remainder of the paper is organized as follows: Section 2 defines some challenges for single-trial detection with an asynchronous stream of stimuli, we define feature extraction methods that are shift invariant in Section 3. The experimental protocol is then detailed in Section 4. The signal processing methods are described in Section 5. Finally, the results are presented and discussed in Sections 6 and 7.

## 2. Challenges for ERP single-trial detection

The detection of single-trial with unknown stimulus onsets implies several challenges. When processing a stream of images without the knowledge of the key stimulus onsets, the first issue is to determine the jitter that can be accepted by the ERP detection procedure. Even when the stimuli are time locked, the amplitude and the latency of the ERP can change over time during an experiment and across sessions. The classifier is therefore able to generalize to some extent the variability in both the amplitude and the latency. However, the classifier performance should be measured to estimate to what extent the classifier is able to detect ERPs that are shifted 100 ms or more. An important shift in the ERP could impair the detection. This observation is critical for most of the current classifiers that are used for ERP detection. Indeed, their architecture is usually shallow, i.e., they are mainly composed of a single linear classifier [24], as opposed to multi-classifier systems [19] and/or non-linear classifiers [5]. With a single linear classifier with the EEG signal as direct input features, a drop of performance can be expected with an important shift of the data, as a linear classifier does not have a high-level representation of the ERP.

With a practical case, if a real time system detects ERPs every 200 ms, it is possible that the system would not be able to detect an ERP that occurs between two detections. With detection occurring every 100 ms, it would be easier to be aligned in time to the evoked potential. However, it would be more difficult to detect several consecutive targets. Indeed, the classifier could be invariant to a shift in time, and it could not be possible to distinguish the presence of several targets. According to previous studies in single-trial ERP detection, the best detection techniques have considered an EEG signal that was bandpass filtered with cutoff frequencies at 1 and 10.66 Hz. Moreover, the sampling frequency used by the winning team of the MLSP 2010 after downsampling was set to 32 Hz [28]. This means that each sampling point that is used as input in the classifier corresponds to 31.25 ms. As the different components of the ERP, e.g., the P300, corresponds to relatively slow waves, a relative tolerance with low shift in time can be expected. Nevertheless, there exists a dilemma: if the ERP detection is tolerant or invariant to some shifts in time then on one hand it will be possible to detect ERPs that are not precisely time-locked, and on the other hand it will be easier to miss the presence of several consecutive targets. To avoid the latter option, we have considered a low target probability during the experiments.

## 3. Related works

Three main strategies are typically proposed for creating pattern recognition systems that allow the detection of signals that can be delayed in time. The first one consists of removing the constraint of the variation that can occur across signals. The goal is to transfer the problem back to the case where every signal is aligned on the same base. In this case, the signals should be aligned to compensate possible shifts by considering some reference features. The second strategy

Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015

is the most commonly used approach. The goal is to extract a set of descriptors, which are invariant to the shift. This approach is often used for images to deal with the multi-orientation invariance after transforming images in polar coordinate [1,29]. Techniques that can be used for the detection of shift invariant ERP include methods such as Fourier descriptor moments, Zernike invariant moments [23,47–49]. Depending on the problem, particular features can be extracted. In [36], the angular information of the external and internal border points of the characters is used to extract rotation invariant features. In the third solution, the classifier does not consider the shift that may occur across signals. The constraint is directly absorbed by the classifier [14]. In fact, machine learning with deep architecture can deal with large deformation of the inputs [26]. In addition, the approach depends on the training database that can be used. When the training database contains only signals without the information related to the shift, it is harder to create a prototype that represents only "aligned" signals.

If $r(t)$ is a shifted version of $s(t)$ so that

$$r(t) = s(t - \Delta) \tag{1}$$

where $\Delta$ is an arbitrary delay in signal, a shift-invariant operator $S()$ satisfies the following equation

$$S(r(t)) = S(s(t)) \tag{2}$$

The operator $S$ can allow the extraction of features, in this case the extracted features of $S(r(t))$ and $S(s(t))$ will be the same. The operator $S$ can correspond to a classifier: the output class of $S(r(t))$ will be the same as $S(s(t))$. For ERP corresponding to the presentation of a visual stimulus in an oddball task, we expect to obtain an N2 component (a negative deflection in the EEG waveform at about 200 ms after a stimulus presentation), and a P3 component (a positive deflection in the EEG waveform at about 300 ms after a stimulus presentation). These components can be considered as points to align the EEG signal. An approach to align ERP responses can be based on the detection of a positive or negative waveform.

### 3.1. Shift-invariant based on the DFT

A common used shift-invariant operator is the power spectrum, which corresponds to the Fourier transform of the second-order cumulant, i.e., the autocorrelation. The Discrete Fourier Transform (DFT) of $s$ and $r$ are defined by:

$$F_s(k) = \sum_{t=0}^{t=N-1} s(t) e^{-\frac{2\pi j}{N} tk} \tag{3}$$

$$F_r(k) = \sum_{t=0}^{t=N-1} s(t) e^{-\frac{2\pi j}{N} (t-\Delta)k} \tag{4}$$

where $j = \sqrt{-1}$, and $N$ is the number of points in the signal. Then we have:

$$|F_s(k)| = |F_r(k)| \tag{5}$$

As being invariant to circular shift in time, it means that two shifted signals with the same ERP components will have similar features. The ERP in each signal needs to be fully contained within the signal. The difference will be due to the difference at the beginning or the end of one of the signals.

### 3.2. Shift-invariant based on the bispectrum

A strong disadvantage of the second-order statistics (second-order correlation) is the loss of the information contained in the phase of the signal. Hence, the shift invariant features that are extracted from the power spectrum will have a discriminant power depending on the phase information of the ERP. If the phase is important, it is likely that the power spectrum will not be enough. To solve this issue, we propose to consider higher-order correlations, or cumulants, (order $>2$) of a signal in order to extract information that can be related to the phase or presence of non-linearities. Higher-order spectra are defined as the multidimensional Fourier transforms of higher-order cumulants. Cumulants of Gaussian processes, of order superior to two, are identically zero. Therefore, higher-order spectra are high signal-to-noise-ratio domains, where system identification or signal reconstruction can be performed.

The bispectrum of a signal $s$ is the two-dimensional DFT of its triple-correlation [35]. The bispectrum has been used in several studies for the creation of shift-invariant features [9,10,34,43], with applications in EEG [2]. By considering the application of the FFT on $s$, the bispectrum is defined by:

$$B_s(k, l) = F_s(k) \cdot F_s(l) \cdot F_s^*(o) \tag{6}$$

where $o = (k + l) \bmod N$, $0 \leq k \leq l \leq N/2$, $k + l < N$, and $F_s^*(k)$ is the conjugate of $F_s(k)$. Before computing the bispectrum, we remove from $F_s$ the first value that corresponds to the mean.

The shift-invariance of the bispectrum is established from the definition above and property (1) with simple manipulations:

$$B_r(k, l) = F_r(k) \cdot F_r(l) \cdot F_r^*(o) \tag{7}$$

$$= F_s(k) e^{j\frac{2\pi}{N}k\Delta} \cdot F_s(l) e^{j\frac{2\pi}{N}l\Delta} F_s(o) e^{-j\frac{2\pi}{N}o\Delta} \tag{8}$$

$$= F_s(k) \cdot F_s(l) \cdot F_s^*(o) e^{j\frac{2\pi}{N}(k+l-o)\Delta} \tag{9}$$

$$= F_s(k) \cdot F_s(l) \cdot F_s^*(o) \tag{10}$$

### 3.3. Local shift-invariant distance

In handwritten character recognition, pattern matching distances are often used with KNN classifiers to increase the detection of shapes with geometric deformations [22]. We propose in this section a distance that takes into account the variability that can happen when a shift occurs in the signal, or when the stimulus onsets contain a jitter. We consider two patterns $P$ and $Q$ of size $N_t \times N_f$, where $N_t$ is the number of points in the time dimension, and $N_f$ is the number of points in the spatial dimension. The distance between $P$ and $Q$ is defined as follows:

$$d(P, Q)_w = \min_{-w \leq k \leq w} \sum_{i_1=1+w}^{N_t-w} \sum_{i_2=1}^{N_f} (P(i_1 + k, i_2) - Q(i_1, i_2))^2 \tag{11}$$

The distance includes the parameter $w$ that determines the maximum shift that can occur to the left or the right in the sampling points.

## 4. Experimental protocol

### 4.1. Procedure and design

Ten healthy subjects (age $= 25.5 \pm 4.4$ years old, three females) participated to a P300 speller experiment, which is the most common paradigm in BCI based on ERP detection. Each subject had to spell a total of 40 characters, and the experiments were carried out sequentially. Each subject observed the same sequence of characters. The matrix of the P300 speller was $6 \times 6$ and displayed on a 27 in. LCD screen with a brightness of 375 cd/m$^2$. Subjects were sitting on a chair at about 60 cm from the screen, in a non shielded room. The stimulus onset asynchrony (SOA) was set to 133 ms and the inter-stimulus interval was 66 ms. In the regular P300 paradigm, the flashes are randomly intensified by a block of 12 flashes (6 rows + 6 columns). As the target is going to be intensified two times during a block of 12 flashes, two consecutive flashes can occur on the target. In such a situation, a user may fail to detect a second salient target occurring in succession if it is presented between 200 and 500 ms after the first one. The expected P300 of the second flash can have a low amplitude, which can impair its detection. The target epochs with a target to target interval

Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015

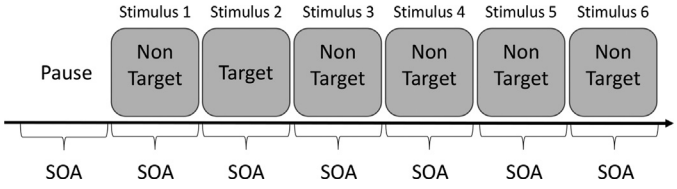



**Fig. 1.** Example of a sequence of stimuli.

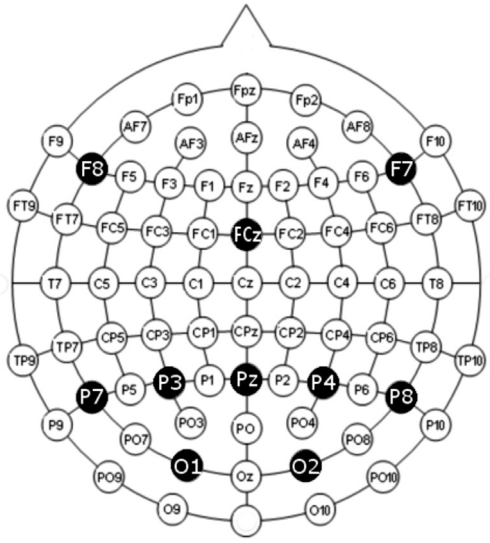



**Fig. 2.** Electrode placement.

of about 200 ms are characterized by a severely reduced classification accuracy that is close the chance level [32]. To optimize the sequence of visual stimuli, we separate the block of flashes: first the row, then the column. In addition, we consider a pause between the block of rows and the block of columns, and between each repetition, to avoid two consecutive flashes in a too short time (see Fig. 1). These pauses lead to a slower speller: 14 SOA instead of 12 for each repetition. During the experiments, the duration of the pause is equal to 1 SOA. Hence, if we consider an SOA of 133 ms, there will be in average more than 0.9 s between two visual stimuli on the target.

### 4.2. Signal acquisition

Electrodes were placed according to a subsampled version of the 10–10 system [45]. The EEG signal was recorded on $O_1$, $O_2$, $P_3$, $P_4$, $P_7$, $P_8$, $P_Z$ and $FC_Z$. $F7$ and $F8$ were dedicated to the ground and the reference, respectively (see Fig. 2). The choice of the electrodes was based on previous studies [8]. The amplifier was a FirstAmp (Brain Products GmbH) with a sampling frequency of 2 kHz. To match real application, all the trials were considered and no specific artifact rejection technique was applied.

## 5. Classification

### 5.1. Signal processing

Before the classification step, a set of features was determined for what best discriminates the ERP to targets from the ERP to non targets. The experimental protocol suggested the presence of a P300 in the ERP of each target. Thus, the signal could be analyzed between the beginning of the visual stimulus and less than 1 s after its beginning. The signal was first bandpassed filtered (Butterworth filter of order 4), with zero-phase distortion, with cutoff frequencies at 1 and 10.66 Hz. Then the signal was downsampled to obtain a signal at a sampling rate equivalent to 25 Hz. For the following steps, we used the observed signal over 640 ms after the start of a visual stimulus, which corresponds to 16 sampling points ($N_e = 16$).

The next step consisted of enhancing the relevant signal by using spatial filters. Let us denote by $U \in \mathbb{R}^{N_s \times N_f}$, the spatial filters, where $N_s$ is the total number of sensors and $N_f$ is the number of spatial filters. The signal after spatial filtering is defined by $X_{\text{filt}} = XU$ where $X \in \mathbb{R}^{N_t \times N_s}$ is the recorded signal, $N_t$ is the number of sampling points in the recorded signal that is considered for determining $U$.

For spatial filtering, the expected ERP is considered as stable over time. Although, the latency and amplitude of the P300 may vary over time for a given task, a spatially stationary waveform of the ERP is considered. With this hypothesis, a single set of spatial filters was applied across the whole signal. We consider here the xDAWN algorithm [42]. This method has been already successfully applied in BCI for P300 detection in the P300 speller paradigm [41]. An algebraic model of the recorded signals $X$ is composed of three terms: the responses on targets ($D_1A_1$), a response common to all stimuli, i.e., targets and non-targets confound ($D_2A_2$) and the residual noise ($H$)

$$X = D_1A_1 + D_2A_2 + H. \tag{12}$$

where $D_1$ and $D_2$ are two real Toeplitz matrices of size $N_t \times N_1$ and $N_t \times N_2$ respectively. $D_1$ has its first column elements set to zero except for those that correspond to a target onset, which are represented with a value equal to one. For $D_2$, its first column elements are set to zero except for those that correspond to stimulus onset. $N_1$ and $N_2$ are the number of sampling points representing the evoked potentials on the target, like the P300 response, and superimposed evoked potentials common to target and non-target, respectively. $H$ is a real matrix of size $N_t \times N_s$.

The goal of applying spatial filters $U \in \mathbb{R}^{N_s \times N_f}$ is to enhance the signal to signal-plus-noise ratio (SSNR) of the enhanced P300 responses ($D_1A_1U$), where $N_f$ is the number of spatial filters

$$XU = D_1A_1U + D_2A_2U + HU. \tag{13}$$

The SSNR in relation to the spatial filters is defined by:

$$\text{SSNR}(U) = \frac{Tr(U^T\hat{A}_1^T D_1^T D_1 \hat{A}_1 U)}{Tr(U^T X^T X U)} \tag{14}$$

where $\hat{A_1}$ corresponds to the least mean square estimation of $A_1$ :

$$\hat{A} = \begin{bmatrix} \hat{A_1} \\ \hat{A_2} \end{bmatrix} = ([D_1; D_2]^T[D_1; D_2])^{-1}[D_1; D_2]^T X \tag{15}$$

where $[D_1; D_2]$ is a matrix of size $N_t \times (N_1 + N_2)$ obtained by concatenation of $D_1$ and $D_2$.

Spatial filters are obtained through the Rayleigh quotient by maximizing the SSNR after two QR decompositions and a singular value decomposition [42]:

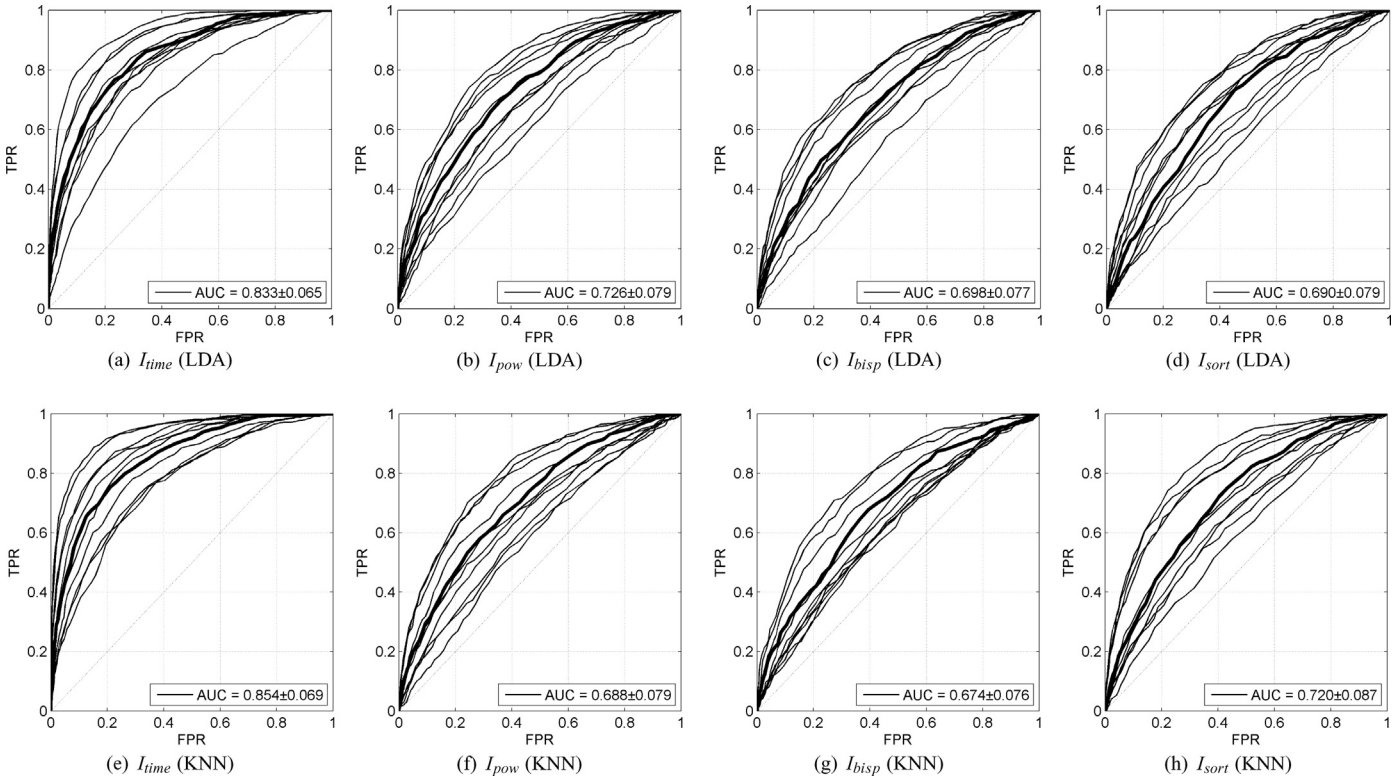
$$\hat{U} = \text{argmax}_U \, \text{SSNR}(U). \tag{16}$$

### 5.2. Input features

For the classifier input, four spatial filters ($N_f = 4$) were used. This number was chosen in relation to prior studies with P300 based BCIs [8]. We consider four sets of input features:

- Time domain ($I_{\text{time}}$). The number of features is $N_f \cdot N_e = 64$.
- Power Spectrum ($I_{\text{pow}}$). The number of features is $N_f \cdot 9 = 36$. Given the size of $N_e = 16$, we apply the Fast Fourier Transform ($N = 16$) and we consider for each time-course signals across spatial filters the amplitude of the 9 first points of its results.
- Bispectrum ($I_{\text{bisp}}$). The number of features is $N_f \cdot 10$. For each component, we only select a subselection of features from the bispectrum.

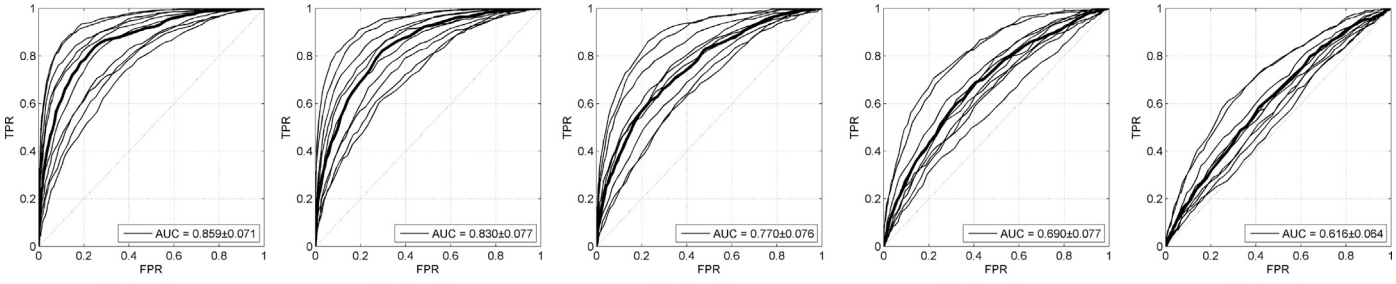
Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015




(a) $I_{time}$ (LDA) (b) $I_{pow}$ (LDA) (c) $I_{bisp}$ (LDA) (d) $I_{sort}$ (LDA)

(e) $I_{time}$ (KNN) (f) $I_{pow}$ (KNN) (g) $I_{bisp}$ (KNN) (h) $I_{sort}$ (KNN)

**Fig. 3.** ROC curves for each subject and each method. The evaluation is based on the correct stimulus onsets, with LDA and KNN as classifiers.




(a) (w=1) (b) (w=2) (c) (w=3) (d) (w=4) (e) (w=5)

**Fig. 4.** ROC curves for each subject and each method. The evaluation is based on the correct stimulus onsets, with the shift invariant distance and KNN.

- Sorted values in time domain ($I_{\text{sort}}$). The number of features is $N_f \cdot N_e = 64$. This simple method allows to create a basic set of features that is shift invariant.

For the shift invariant distance, we only use $I_{\text{time}}$.

### 5.3. Classifier

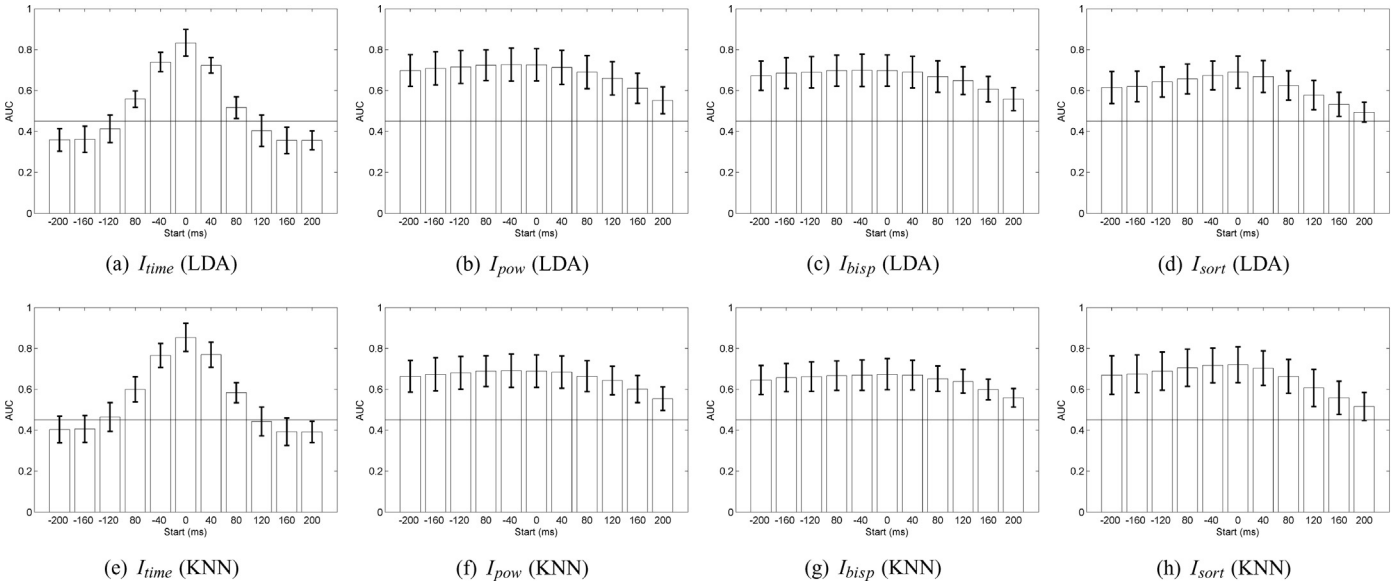
The input vector of the classifier is obtained by the concatenation of the $N_f$ time-course signals across spatial filters. The linear discriminant analysis (LDA) and $k$-nearest neighbor (KNN) ($k = 20$) classifier are considered for the detection. For the evaluation of the classifier, we provide the results obtained after a two folds cross-validation, each fold corresponds to an experimental session. The evaluation of the classification tasks is assessed by the area under the curve (AUC) of the ROC (receiver operating characteristic) curves. ROC curves allow analyzing and visualizing the classifier performance. As the classifier output does not produce a class decision, but a confidence measure, we consider non-parametric ROC curves for the classifier performance estimation [13].

## 6. Results

### 6.1. Single-trial detection

In this section, we report the performance for single-trial detection when the stimulus onsets are known for both training and the test. The ROC curves for each subject and each set of input features is presented in Figs. 3 and 4. The curves in bold represent the estimated ROC curve corresponding to the average AUC across subjects. The AUC across subjects are $0.833 \pm 0.064$ for $I_{\text{time}}$, $0.726 \pm 0.079$ for $I_{\text{pow}}$, $0.698 \pm 0.077$ for $I_{\text{bisp}}$, and for $0.690 \pm 0.079$ $I_{\text{sort}}$ with an LDA classifier. The AUC is $0.854 \pm 0.069$ for $I_{\text{time}}$, $0.688 \pm 0.079$ for $I_{\text{pow}}$, $0.674 \pm 0.076$ for $I_{\text{bisp}}$, and for $0.720 \pm 0.087$ $I_{\text{sort}}$ with a KNN classifier. As expected, a drop of performance is observed from $I_{\text{time}}$ to $I_{\text{pow}}$, $I_{\text{bisp}}$ and $I_{\text{sort}}$ by the addition of the shift invariance constraint. A repeated measures analysis of variance (ANOVA) confirms the difference across methods, ($F = 35.94$, $p < 10e{-}10$) for LDA, and ($F = 84$, $p < 10e{-}10$) for KNN. Post-hoc $t$-tests after a Bonferoni correction indicate that the input features with the power spectrum involve a clear drop of performance compared to the input features in the time domain ($t_9 = 4.86$,

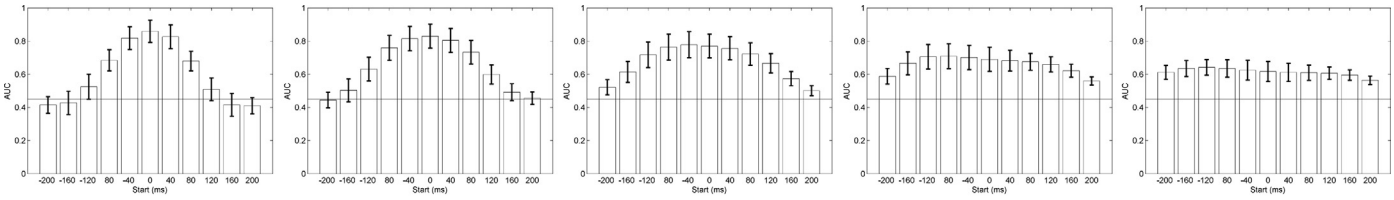
Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015




(a) $I_{time}$ (LDA) (b) $I_{pow}$ (LDA) (c) $I_{bisp}$ (LDA) (d) $I_{sort}$ (LDA)

(e) $I_{time}$ (KNN) (f) $I_{pow}$ (KNN) (g) $I_{bisp}$ (KNN) (h) $I_{sort}$ (KNN)

**Fig. 5.** AUC for different shifts of the stimulus onsets, from −200 ms to +200 ms after the real stimulus onsets, with LDA and KNN classifiers. The error bars represent the standard deviation across subjects.




(a) w=1 (b) w=2 (c) w=3 (d) w=4 (e) w=5

**Fig. 6.** AUC for different shifts of the stimulus onsets, from −200 ms to +200 ms after the real stimulus onsets, for the shift-invariant distance using KNN. The error bars represent the standard deviation across subjects.

$p < 10e-4$) for LDA, ($t_9 = 10.37$, $p < 10e-6$) for KNN. The same type of comparison shows that the features from the $I_{\text{bisp}}$ and $I_{\text{sort}}$ provides a worse performance than $I_{\text{time}}$. For the comparison across all the methods, the best performance is obtained with KNN and the shift invariant distance with an AUC of $0.859 \pm 0.071$. However, post-hoc $t$-tests, after a Bonferoni correction, show that there is no significant difference between KNN and LDA with inputs in the time domain ($I_{\text{time}}$). Nevertheless, the results indicate that allowing a local shift invariance in the classifier may increase the robustness of the system. The difference of performance across classifiers also shows that LDA provides better performance than KNN with $I_{\text{pow}}$ ($t_9 = 4.06$, $p < 10e-2$), and KNN provides better performance than LDA with $I_{\text{sort}}$ ($t_9 = 3.72$, $p < 10e-2$).

### 6.2. Different stimulus onsets

During the test, we consider different onsets for the stimuli. These onsets correspond to different shifts of the signal. The AUC across subjects is presented in Figs. 5 and 6 for different shifts of the visual stimulus onsets. They correspond to a left and right shift from 1 to 5 time points, which is equivalent to a shift of 40 ms to 200 ms. When we consider the inputs from $I_{\text{time}}$, with a shift to the left or to the right of 80 ms during the test, the performance of the classifier is severely decreased, and the mean AUC across subjects becomes around 0.51, close to a decision given by chance. On the other hand, $I_{\text{pow}}$, $I_{\text{bisp}}$ and $I_{\text{sort}}$ allow keeping a stable performance in spite of the different shifts. Across the ten shifts and the initial onset, the mean AUC across subjects is $0.510 \pm 0.178$ for the $I_{\text{time}}$, $0.684 \pm 0.056$ for the $I_{\text{pow}}$, $0.665 \pm 0.045$ for the $I_{\text{bisp}}$, and for $0.617 \pm 0.061$ $I_{\text{sort}}$ for LDA, and $0.552 \pm 0.174$ for the $I_{\text{time}}$, $0.657 \pm 0.043$ for the $I_{\text{pow}}$, $0.645 \pm 0.036$ for the $I_{\text{bisp}}$, and for $0.657 \pm 0.068$ $I_{\text{sort}}$ for KNN. With the shift invariant distance, the mean AUC is $0.597 \pm 0.181$, $0.642 \pm 0.064$, $0.671 \pm 0.062$, $0.660 \pm 0.058$, and $0.613 \pm 0.022$ for $w$ from 1 to 5. The best approach is the combination of LDA and $I_{\text{pow}}$.

If the distribution of the shifts around the real stimulus onsets follows (±200 ms)a standard normal distribution (standard deviation is equal to one), then the mean AUC across shifts is 0.748 for $I_{\text{time}}$, 0.720 for $I_{\text{pow}}$, 0.694 for $I_{\text{bisp}}$, and 0.675 for $I_{\text{sort}}$, for LDA; 0.780 for $I_{\text{time}}$, 0.686 for $I_{\text{pow}}$, 0.670 for $I_{\text{bisp}}$, and 0.711 for $I_{\text{sort}}$, for KNN. For KNN and the shift invariant distance, the AUC is 0.819, 0.809, 0.765, 0.690, and 0.618 for $w$ from 1 to 5. This analysis shows that for systems where the variation around the exact stimulus onsets follow a standard normal distribution, it is better to use KNN with a shift invariance distance, with parameters that can allow a difference between the onsets up to 80 ms.

This observation highlights the critical impact of the synchronization between the onset corresponding to presentation of the stimulus to the user, and the onset corresponding to the classification step. Such a drop of performance was expected as the main component of the ERP (P100, N200, P300, ...) are separated by 100 ms. A shift of about 80 ms can completely impair the classification performance when the inputs are in the time domain whereas the performance remains stable with input features in the Fourier domain or when an appropriate distance is chosen.

## 7. Discussion

The detection of ERPs has evolved over the two last decades for different types of application thanks to machine learning techniques that have improved the detection of brain responses like ERPs. While

Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015

the first works about ERP have mainly focused on the grand average [16,39], and still provide useful information about ERPs, ERPs have been analyzed with a limited number of trials like in BCI [5,31]. This evolution has continued toward the detection of single-trial evoked responses [44]. For all these types of studies, i.e., from the grand average analysis to single-trial detection, the information of the stimulus onsets has always been part of the prior knowledge for the analyses. The problem of single-trial detection is far from being solved. It needs further investigation, such as the improvements of the paradigms, machine learning, and signal processing techniques. A new trend is to overcome the necessity of knowing the stimulus onsets for proposing applications (clinical and/or commercial) of target detection in dynamic environments with events occurring in real-time.

In recent studies about the detection of ERPs for the P300 speller [5,40] or for RSVP tasks [17], the inputs of the classifier are always considered in the time domain. While this strategy has been successfully applied when the stimulus onsets are known, the absence of the stimulus onsets leads to the creation of different input features sets, which are shift-invariant. In terms of performance, the Fourier domain proposes a good solution for determining shift-invariant features that can be exploited when the stimulus onsets are unknown or imprecise. Yet, the performance in the Fourier domain is lower than in the time domain when the stimulus onsets are known.

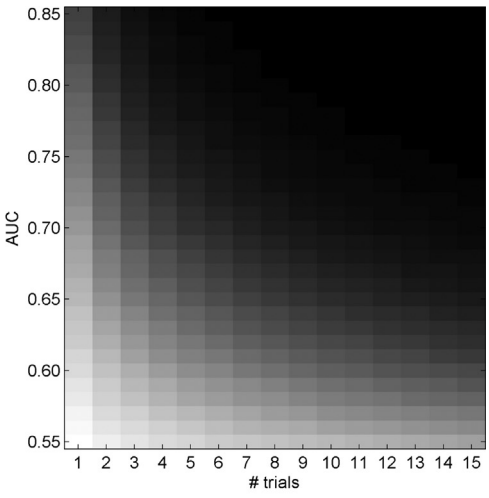
### 7.1. Potential applications

With wireless EEG amplifiers and wireless devices such as smartphone and tablets, it is likely that a perfect synchronization between the EEG acquisition and the device of the user may not be possible. Despite the performance that can be obtained with features in the time domain with a perfect synchronization between the EEG and the stimulus onsets, we have shown that a significant drop of performance can occur with a shift of 40 ms in the stimulus onsets, from AUC $= 0.833$ to about AUC $= 0.723$, with an LDA classifier. In this case, features with the power spectrum, albeit less efficient, may provide a more robust performance with an important jitter. More importantly, we have shown that the use of KNN with a shift invariant distance can limit this drop of performance, the AUC only drops from 0.859 to about 0.820 when there is a shift of 40 ms. This new approach is particularly relevant for wireless EEG amplifiers and relatively inexpensive amplifiers that do not allow precise recording of stimulus onsets.

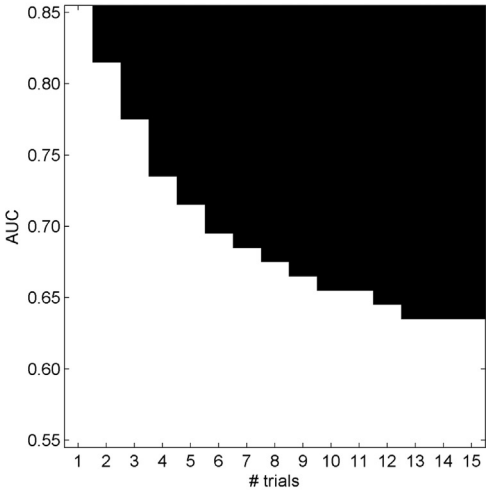
With data acquired from a P300 speller experiment, the training of the classifier was achieved by considering the exact stimulus onsets. The task was similar between the training and test sessions. If single-trial detection should be used for video or in a real environment, a judicious calibration session should be chosen to properly represent the different parameters of the test session, e.g., target probability, stimulus meaning, as these parameters have an influence on the ERPs, e.g., the amplitude of the P300 [16]. A calibration session that does not correctly represent the test session would impair the performance of the classifier.

### 7.2. Limitations

Single-trial detection with shift invariance would show some limits when consecutive targets are presented, and each of them should be detected individually. As the system is invariant to some translation, it would assign a target to different events that are close to the target. This characteristic can be an issue for fast tasks where the detection of an ERP should be precisely assigned to a particular event. Hence, the proposed strategy should only be considered for tasks with a large inter-target interval ($>$1s), where events to detect are rare such as the detection of threats in dynamic real-world and real-time tasks. The main priority of the application should be to detect the presence of an event while allowing a certain time range for its presence, i.e., the system is able to detect the presence of an event, but the precision of its presentation remains imprecise, e.g., $\pm$100 ms.




(a) AUC




(b) AUC $>$ 0.9

**Fig. 7.** AUC as a function of the AUC for single-trial detection and the number of trials. A dark color represents an AUC close to 1 (a), the black color represents AUC values superior to 0.9 (b).

Single-trial detection can be combined across trials with repetitions of the stimuli such as in the P300 speller, or with trials from a group of subjects in cooperative BCIs [6,7]. Fig. 7 depicts the AUC that is obtained with Monte-Carlo simulations by considering a linear combination of the scores coming from several trials. With the AUC obtained from the experiment and by considering a jitter following a normal distribution ($\pm$200 ms), it would be necessary to combine at least 13 trials with $I_{\text{time}}$, 6 trials with $I_{\text{time}}$, 7 trials with $I_{\text{time}}$, to reach a mean AUC of 0.90. This result shows how a difference in the AUC can impact the overall performance.

## 8. Conclusion

Methods including feature sets based on the Fourier domain, and a local invariant distance have been proposed for the detection of single-trial ERP when an important jitter between the stimulus onsets and EEG recordings may occur. This work represents a step toward the detection of single-trial evoked responses in difficult conditions, thus leveraging ERP based BCI applications. The method considers spatial filters based on the maximization of the signal-to-signal plus noise

Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015

ratio that are determined on data with stimulus onsets. After spatial filtering, a set of shift-invariant features is obtained with the Fourier transform and/or the bispectrum, or by considering a shift invariant distance. With this approach, we have shown that it is possible to detect ERPs with a jitter of $\pm 200$ ms. The classifier performance (AUC $> 0.70$) is significantly above chance, hence supporting the conclusion that ERPs can be detected without any knowledge of the precise stimulus onsets. As BCI will become available for healthy people with different wireless devices in a near future, new feature extraction approaches should be considered that include this constraint.

## References

[1] S. Adam, J.M. Ogier, C. Carlon, R. Mullot, J. Labiche, J. Gardes, Symbol and character recognition: application to engineering drawing, Int. J. Doc. Anal. Recogn. 3 (2000) 89–101.
[2] R.J.G. Ajraj, M. Doi, H. Mantzaridis, G.N.C. Kenny, Analysis of the EEG bispectrum, auditory evoked potentials and the EEG power spectrum during repeated transitions from consciousness to unconsciousness, Br. J. Anaesth. 80 (1998) 46–52.
[3] N. Bigdely-Shamlo, A. Vankov, R.R. Ramirez, S. Makeig, Brain activity-based image classification from rapid serial visual presentation, IEEE Trans. Neural Syst. Rehabil. Eng. 16 (5) (2008) 432–441.
[4] H. Cecotti, A self-paced and calibration-less SSVEP based brain-computer interface speller, IEEE Trans. Neural Syst. Rehabil. Eng. 18 (2010) 127–133.
[5] H. Cecotti, A. Gräser, Convolutional neural networks for P300 detection with application to brain-computer interfaces, IEEE Trans. Pattern Anal. Mach. Intell. 33 (3) (2011) 433–445.
[6] H. Cecotti, B. Rivet, Performance estimation of a cooperative brain-computer interface based on the detection of steady-state visual evoked potentials, in: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014a, pp. 1–4.
[7] H. Cecotti, B. Rivet, Subjects and electrodes combination in cooperative brain-computer interface based on event-related potentials, Brain Sci. 4 (2014b) 335–355.
[8] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, J. Mattout, A robust sensor selection method for P300 brain-computer interfaces, J. Neural Eng. 8 (2011) 016001.
[9] V. Chandran, B. Carswell, B. Boashash, S.L. Elgar, Pattern recognition using invariants defined from higher order spectra – 2-d image inputs, IEEE Trans. Image Process. 6 (5) (1997) 703–712.
[10] V. Chandran, S.L. Elgar, Pattern recognition using invariants defined from higher order spectra – one-dimensional inputs, IEEE Trans. Signal Process. 41 (1) (1993) 205–212.
[11] M.M. Chun, C.M. Potter, A two-stage model for multiple target detection in rapid serial visual presentation, J. Exp. Psychol. Hum. Percept. Performance 21 (1) (1995) 109–127.
[12] L. Farwell, E. Donchin, Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials, Electroencephalogr. Clin. Neurophysiol. 70 (1988) 510–523.
[13] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (2006) 861–874.
[14] M. Fukumi, S. Omatu, T. Takeda, T. Kosaka, Rotation invariant neural pattern recognition system with application to coin recognition, IEEE Trans. Neural Networks 3 (1992) 272–279.
[15] A. Gerson, L. Parra, P. Sajda, Cortically-coupled computer vision for rapid image search, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 174–179.
[16] C.J. Gonsalvez, J. Polich, P300 amplitude is determined by target-to-target interval, Psychophysiology 39 (2002) 388–396.
[17] K.E. Hild, M. Kurimo, V.D. Calhoun, The sixth annual MLSP competition, 2010, in: Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2010, pp. 107–111.
[18] S. Hirata, G. Matsuda, A. Ueno, H. Fukushima, K. Fuwa, K. Sugama, K. Kusunoki, M. Tomonaga, K. Hiraki, T. Hasegawa, Brain response to affective pictures in the chimpanzee, Sci. Rep. 3 (2013) 1342.
[19] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, Boosting linear logistic regression for single trial ERP detection in rapid serial visual presentation tasks, in: Proceedings of the IEEE Engineering in Medicine and Biology Society, 2006, pp. 3369–3372.
[20] Y. Huang, D. Erdogmus, S. Mathan, M. Pavel, Large-scale image database triage via EEG evoked responses, in: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2008, pp. 429–432.
[21] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, K.E. Hild, A framework for visual image search using single-trial brain responses, Neurocomputing. 74 (2011) 2041–2051.
[22] D. Keysers, T. Deselaers, C. Gollan, H. Ney, Deformation models for image recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (8) (2007) 1422–1435.
[23] A. Khotanzad, Y.H. Hong, Invariant image recognition by Zernike moments, IEEE Trans. Pattern Anal. Mach. Intell. 12 (5) (1990) 489–497.
[24] D.J. Krusienski, E.W. Sellers, F. Cabestaing, S. Bayoudh, D.J. McFarland, T.M. Vaughan, J.R. Wolpaw, A comparison of classification techniques for the P300 speller, J. Neural Eng. 3 (2006) 299–305.
[25] S.F. Lang, C.A. Nelson, P.F. Collins, Event-related potentials to emotional and neutral stimuli, J. Clin. Exp. Neuropsychol. 12 (6) (1990) 946–958.
[26] H. Larochelle, D. Erhan, A. Courville, J. Bergstra, Y. Bengio, An empirical evaluation of deep architectures on problems with many factors of variation, in: Proceedings of the 24th International Conference on Machine learning, 2007, pp. 473–480.
[27] R. Leeb, D. Friedman, G.R. Müller-Putz, R. Scherer, M. Slater, G. Pfurtscheller, Self-paced (asynchronous) BCI control of a wheelchair in virtual environments: a case study with a tetraplegic, Comput. Intell. Neurosci. (2007).
[28] J.M. Leiva, S.M.M. Martens, MLSP competition, 2010: description of the first place method, in: Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP), 2010, pp. 112–113.
[29] S.X. Liao, M. Pawlak, On image analysis by moments, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 254–266.
[30] S.J. Luck, An Introduction to the Event-Related Potential Technique, The MIT Press, 2005.
[31] J.N. Mak, Y. Arbel, J.W. Minett, L.M. McCane, B. Yuksel, D. Ryan, D. Thompson, L. Bianchi, D. Erdogmus, Optimizing the P300-based braincomputer interface: current status, limitations and future directions, J. Neural Eng. 8 (2011) 025003.
[32] S.M. Martens, N.J. Hill, J. Farquhar, B. Schölkopf, Overlap and refractory effects in a brain-computer interface speller based on the visual P300 event-related potential, J. Neural Eng. 6 (2) (2009) 026003.
[33] J. Millán, J. Mouriño, Asynchronous BCI and local neural classifiers: an overview of the adaptive brain interface project, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 159–161.
[34] R.M.P. Negrinho, P.M.Q. Aguiar, Shape representation via elementary symmetric polynomials: a complete invariant inspired by the bispectrum, in: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2013, pp. 1–5.
[35] C.L. Nikias, A.P. Petropulu, Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework, PTR Prentice Hall, Englewood Cliffs, NJ, 1993.
[36] U. Pal, F. Kimura, K. Roy, T. Pal, Recognition of English multi-oriented characters, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2006, pp. 873–876.
[37] R. Panicker, S. Puthusserypady, Y. Sun, An asynchronous P300 BCI with SSVEP-based control state detection, IEEE Trans. Biomed. Eng. 58 (6) (2011) 1781–1788.
[38] L.C. Parra, C. Christoforou, A.D. Gerson, M. Dyrholm, A. Luo, M. Wagner, M.G. Philiastides, P. Sajda, Spatio-temporal linear decoding of brain state: Application to performance augmentation in high-throughput tasks, IEEE Signal Process. Mag. 25 (1) (2008) 95–115.
[39] J. Polich, Updating P300: an integrative theory of P3a and P3b, Clin. Neurophysiol. 118 (2007) 2128–2148.
[40] A. Rakotomamonjy, V. Guigue, BCI competition III : dataset II – ensemble of SVMs for BCI P300 speller, IEEE Trans. Biomed. Eng. 55 (3) (2008) 1147–1154.
[41] B. Rivet, H. Cecotti, E. Maby, J. Mattout, Impact of spatial filters during sensor selection in a visual P300 brain-computer interface, Brain Topogr. 12 (1) (2012) 55–63.
[42] B. Rivet, A. Souloumiac, V. Attina, G. Gibert, xDAWN algorithm to enhance evoked potentials: application to brain-computer interface, IEEE Trans. Biomed. Eng. 56 (8) (2009) 2035–2043.
[43] B. Sadler, G.B. Giannakis, Shift and rotation invariant object reconstruction using the bispectrum, J. Opt. Soc. Am. A 9 (1992) 57–69.
[44] P. Sajda, A. Gerson, L. Parra, High-throughput image search via single-trial event detection in a rapid serial visual presentation task, in: Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering, 2003, pp. 7–10.
[45] F. Sharbrough, G. Chatrian, R.P. Lesser, et al., Guidelines for Standard Electrode Position Nomenclature, American EEG Society, Bloomfield, IL, 1990.
[46] J. Touryan, L. Gibson, H.J. Horne, P. Weber, Real-time measurement of face recognition in rapid serial visual representation, Front. Psychol. 2 (42) (2011) 1–8.
[47] D.M. Tsai, H.C. Chiang, Rotation-invariant pattern matching using wavelet decomposition, Pattern Recognit. Lett. 23 (2002) 191–201.
[48] W.H. Wong, W.C. Siu, K.M. Lam, Generation of moment invariants and their uses for character recognition, Pattern Recognit. Lett. 16 (2) (1995) 115–123.
[49] T.N. Yang, S.D. Wang, A rotation invariant printed Chinese character recognition system source, Pattern Recognit. Lett. 22 (2) (2001) 85–95.

Please cite this article as: H. Cecotti, Toward shift invariant detection of event-related potentials in non-invasive brain-computer interface, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.01.015